Massive Molecular Parallel Evolution of the *HSP90AA1* Gene between High-elevation Anurans

Hong JIN[1], Bin LU[1] and Jinzhong FU[1,2*]

[1] *Chengdu Institute of Biology*, *Chinese Academy of Sciences*, *Chengdu* 610041, *China*
[2] *Department of Integrative Biology*, *University of Guelph*, *Guelph*, *Ontario N1G*2*W*1, *Canada*

Abstract *HSP90AA1* is part of the heat shock protein 90 gene family and has important functions against heat stress. We report a case of molecular level parallel evolution of the *HSP90AA1* gene in high elevation amphibians. *HSP90AA1* gene sequences of four high-elevation anurans, *Bufo gargarizans, Nanorana parkeri, Rana kukunoris,* and *Scutiger boulengeri*, were compared along with five of their low-elevation relatives. A total of 16 amino-acid sites were identified as parallel evolution between *N. parkeri* and *R. kukunoris*. We generated both model based (Zhang and Kumar's test) and empirical data based (parallel/divergence plotting) null distributions for non-parallel evolution, and both methods clearly determined that the observed number of parallel substitutions were significantly more than the null expectation. Furthermore, on the *HSP90AA1* gene tree, *N. parkeri* and *R. kukunoris* formed a strongly supported clade that was away from their respective relatives. This study provides a clear case of molecular parallel evolution, which may have significant implications in understanding the genetic mechanisms of high-elevation adaptation.

Keywords molecular parallel evolution, high-elevation, case study, amphibian, *HSP90AA1* gene

1. Introduction

Parallel and convergent evolution is an important component of the evolutionary process and there are numerous well-documented examples, such as the wings of birds and bats. Nevertheless, at molecular level, they are generally considered uncommon (e.g. Thomas and Hahn, 2015; Zou and Zhang, 2015a). So far, there are only a few well-established cases of molecular parallel and convergent evolution (e.g. Li *et al*., 2010; Shen *et al*., 2010; Shen *et al*., 2012; Liu *et al*., 2014). The best-studied case is probably the echolocation related genes between echolocating bats and whales. For example, there are strong signals of molecular parallel and convergent evolution of the *prestin* gene (Li *et al*., 2010; Shen *et al*., 2012; Liu *et al*., 2014). The *rhodopsin* (*RH1*) gene in bats for dim light is perhaps another clear case (Shen *et al*., 2010). Furthermore, molecular parallel and

convergent evolution is often considered an indicator of adaptive evolution (e.g. Castoe *et al*., 2009; Liu *et al*., 2014; but see Zou and Zhang, 2015b). Although rare, these evolutionary patterns have significant biological implications (Rokas and Carroll, 2008; Stern, 2013). For example, gene trees derived from genes with parallel or convergent evolution may deviate significantly from the species tree (e.g. Castoe *et al*., 2009; Li *et al*., 2010; Shen *et al*., 2012). There are several controversial issues related to molecular parallel and convergent evolution. For example, rather than rare, Parker *et al*. (2013) reported widespread molecular parallel and convergent evolution between echolocating bats and whales in their genomes. Furthermore, methods of detecting molecular parallel and convergent evolution are highly contentious. Most arguments are centered by how to construct a null distribution for evolution without parallel and convergent evolution (Zhang and Kumar, 1997; Castoe *et al*., 2009; Parker *et al*., 2013; Thomas and Hahn, 2015; Zou and Zhang, 2015a,b).

High-elevation adaptation is a classic example of adaptive evolution and recent works have identified

several genes that are directly involved in the adaptation process (e.g. Simonson *et al.*, 2010; Scott *et al.*, 2011; Yang *et al.*, 2012). Amphibians are no strangers to high-elevation environments. For example, more than 14 amphibians have been recorded at elevations above 3 000 m at the Tibetan Plateau region, and many of them have reduced tympanum and stapes, which are often considered adaption to high elevation environments (Zhao and Yang, 1997; Li *et al.*, 2010). Our recent work detected significant molecular parallel and convergent evolution of the *MYBPC2* gene between three high-elevation amphibian species, *Rana kukunoris*, *Nanorana pleskei*, and *Bufo gargarizans minshanicus*, which may have contributed to their success at high elevation environments (Yang *et al.*, 2017). During the analysis, we also detected signals of probable parallel evolution of another gene, *HSP90AA1*, which may deserve further examination.

HSP90AA1 (ENSG00000080824) is the Class A Member 1 of the Heat Shock Protein 90 Alpha Family. The gene family encodes a set of chaperone proteins that assist other proteins to fold properly, stabilize proteins against heat stress, and aid in protein degradation. The *HSP90AA1* gene encodes the heat shock protein 90α (Hsp90α), a stress-inducible molecular chaperone protein. The protein promotes the maturation, structural maintenance, and proper regulation of specific target proteins involved in cell cycle control and signal transduction (NCBI, 2018). At high elevations, although ambient temperature is generally lower than that at low elevations, it has high UV radiation and large diurnal temperature variation (Cheviron and Brumfield, 2012). It is conceivable that this gene family may be involved in adaptation to high elevations, and a mutation that might improve the ability to deal with heat stress would be beneficial in such environments.

In this study, we compared the *HSP90AA1* gene of four high-elevation amphibian species and evaluated potential parallel and convergent evolution among them. Data were collected both online and by experiments. Multiple detecting methods were applied.

2. Materials and Methods

2.1. Species examined We compared four sets of species. For each set, one high-elevation species and at least one low elevation close relative were included. High-elevation species were defined as species predominately distributed at elevations higher than 3 000 m above sea level. The four high-elevation species were: *Bufo gargarizans, Nanorana parkeri, Rana kukunoris,* and *Scutiger boulengeri*. The latter three are exclusively distributed at high altitudes (3 000 m or higher) and the samples of *Bufo gargarizans* were collected from a high-elevation population (Zoige, 3 340 m). Based on established phylogeny and availability of data, five of their low-elevation close relatives, *Bufo viridis, Odorrana margaretae, Oreolalax popei, Quasipaa spinosa,* and *Rana chensinensis* were selected. They phylogenetic relationships were summarized based on current anuran phylogenies (e.g. Duellman and Trueb, 1994; Pyron and Wiens, 2011), and are depicted in Figure 1A.

Sequence data of the *HSP90AA1* gene of two anurans (*Xenopus tropicalis*, *Nanorana parkeri*) were obtained from previously published genomic data (Hellsten *et al.*, 2010; Sun *et al.*, 2015). All others were extracted from transcriptome data: five were from previous published works, *Bufo gargarizans* (Yang *et al.*, 2016), *Bufo viridis* (Gerchen *et al.*, 2016)*, Odorrana margaretae* (Qiao *et al.*, 2013)*, Rana chensinensis* (Yang *et al.*, 2012)*, R. kukunoris* (Yang *et al.*, 2012), and three, *Quasipaa spinosa, Oreolalax popei, Scutiger boulengeri,* were obtained in this study. A mixture of heart, liver, and muscle tissues from four individuals of each species were used for mRNA extraction. Paired-end sequencing was conducted on an Illumina HiSeq 2000 platform. Standard bioinformatic analysis was conducted and detailed procedures were provided in Yang *et al.* (2012). There are two known variants of transcripts for this gene (NCBI), and we used the longer one to maximize the information content and to assume homology.

2.2. Detecting parallel and convergent evolution Methods of detecting molecular parallel and convergent evolution are controversial, and the available methods involve different ways of constructing the null distribution (Thomas and Hahn, 2015; Zou and Zhang, 2015a,b). We employed three methods involving both model based null distribution and empirical null distribution, and by so doing we strived to avoid false positives introduced by other evolutionary processes.

We first used Zhang and Kumar's test (Zhang and Kumar, 1997) to screen for signals of any parallel and/or convergent amino-acid substitutions between the four high-elevation species. This test generates a null distribution based on a substitution model and tests whether the observed parallel and convergent amino-acid substitutions are significantly more than expected by random chance (Zhang and Kumar, 1997). The amino-acid substitution model Poisson was used.

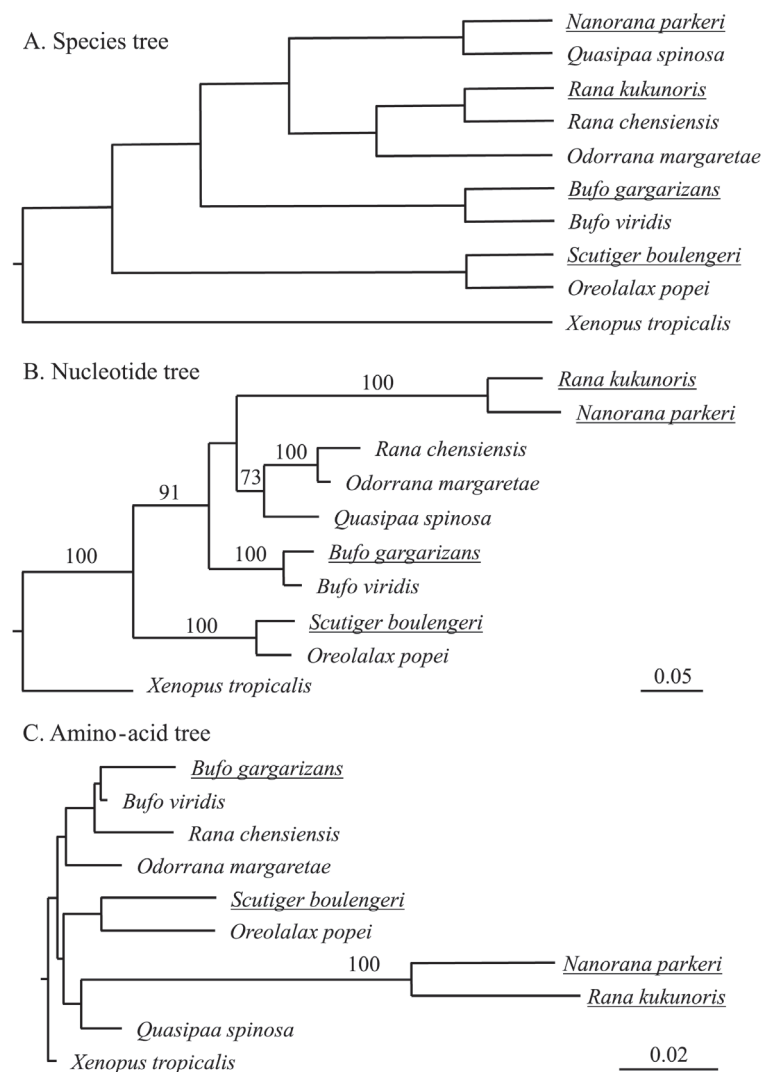We also established an empirical null distribution by

Figure 1 Species tree of ten amphibian species and the *HSP90AA1* gene tree inferred from maximum likelihood analysis. Numbers above branches are bootstrap values and only values greater than 70 are presented. Species with underline are high-elevations species (> 3 000 m). A. A composite species tree from published literature. B. Gene tree derived from nucleotide sequences. C. Gene tree derived from amino-acid sequences.

plotting the pairwise parallel/convergent substitutions against divergent substitutions using R (R Core Team, 2013). The amount of parallel/convergent substitutions by chance is expected to be proportional to the amount of divergent substitutions, and therefore the best-fit line would represent an empirical null distribution (Castoe *et al.*, 2009; Thomas and Hahn, 2015). Ancestral amino-acid sequences for all interior nodes were first reconstructed using a Bayesian method implemented in CODEML (in PAML package; Yang, 2007). Reconstructed sites with posterior probability less than 95% were discarded from analysis, as its state was deemed unreliable. Numbers of divergent and parallel/convergent amino-acid substitutions between all pairs were counted and plotted. We defined divergent substitution as changes occurring on

both branches and having different end states between the two branches. Parallel substitution was defined as changes occurring on both branches and having the same start and end states between the two branches while convergent substitution has different start states but same end state.

Finally, we compared the *HSP90AA1* gene tree to a species tree (Figure 1A; Duellman and Trueb, 1994; Pyron and Wiens, 2011). As several previous studies have discovered (e.g. Castoe *et al.*, 2009; Li *et al.*, 2010; Shen *et al.*, 2012; Yang *et al.*, 2017), strong parallel and convergent evolution could group distantly related species together on a gene tree, and therefore such comparison can be used to detect parallel and convergent evolution. Gene trees were constructed using maximum likelihood (ML) method separately for amino-acid sequences and

nucleotide sequences. *Xenopus tropicalis* was used as outgroup. All sequences were first aligned using ClustalW method, and the best-fit models, WAG+G for amino-acid sequences and GTR+G for nucleotide sequences, were selected based on the Bayesian information criterion. A non-parametric bootstrap was conducted with 1000 replicates. All analyses were carried out using RAxML (Stamatakis, 2014).

2.3. Detecting positive selection Molecular parallel and convergent evolution is often considered an indicator of adaptive evolution (e.g. Castoe *et al.*, 2009; Liu *et al.*, 2014), and therefore we conducted a positive selection test. The branch-site model (test 2 in Zhang *et al.*, 2005) implemented in CODEML was used. Each of the four high-elevation species (Figure 1A) was separately set as the foreground branch and all other branches as background. A likelihood ratio test (LRT) was used to compare the model assuming positive selection to a null model with neutral evolution. If the *P* value of the LRT test was less than 0.05, positive selection may have occurred on the foreground branch.

3. Results

The final alignment of the *HSP90AA1* gene produced a total of 786 base pairs for each species, which translated into 262 amino-acid residues (Figure 2). No gap was introduced.

Zhang and Kumar's test detected significantly more parallel substitutions than expected by random chance between four pairs of high elevation species, including between *R. kukunoris* and *N. parkeri* (*P* < 0.0001), between *R. kukunoris* and *B. gargarizans* (*P* = 0.0069), between *R. kukunoris* and *S. boulengeri* (*P* = 0.0074), and between *N. parkeri* and *S. boulengeri* (*P* = 0.0038).

A total of 94 pairwise comparisons for all terminal and interior branch pairs were conducted and the numbers of parallel/convergent substitutions and divergent substitutions were significantly correlated (Pearson's *r* = 0.0621, *P* = 0.0088; Figure 3), although it was a low correlation. The low *r* value was likely due to most of comparisons having zero parallel/convergent sites. The best-fit line served as an empirical null expectation and two outliers were identified. One was between *R. kukunoris* and *N. parkeri,* for which a total of 16 parallel substitutions were identified (Figure 3). The second one is between *R. kukunoris* and *B. gargarizans*, for which one parallel substitution was identified (Figure 3). No convergent site between *R. kukunoris* and *N. pleskei* or between *R. kukunoris* and *B. gargarizans* was detected.

The gene tree based on nucleotide sequences (Figure 1B) was very similar to the species tree (Figure 1A). However, two distantly related high-elevation species, *Rana kukunoris* and *Nanorana parkeri*, were grouped together and the association was well supported by the bootstrap analysis (bootstrap value = 100). All other relationships were the same as the species tree. The amino-acid gene tree received low nodal support, and only one node received bootstrap values greater than 70% (Figure 1C). This was likely due to the limited number of informative sites. Nevertheless, *R. kukunoris,* and *N. parkeri* were grouped together and the clade received a bootstrap value of 100. Evidently, strong parallel evolution between the two species drew them together, despite the fact that they are distantly related and belong to two different families (families Ranidae and Dicroglossidae, respectively).

We did not detect any significant signal of positive selection. None of the four high-elevation lineages had dN/dS ratios significantly greater than one (*chi²* < 1, *P* > 0.05).

4. Discussion

There is a strong pattern of parallel evolution of the *HSP90AA1* gene between *Rana kukunoris* and *Nanorana parkeri*. All three methods, which involved differently constructed null distributions, unambiguously detected significant parallel evolution than random expectations. A total of 16 amino-acid sites, among 262, appear to be evolving in a parallel fashion (Figure 2). Interestingly, no convergent site was identified. Signals of parallel evolution were also detected between other pairs of high-elevation species, such as between *R. kukunoris* and *B. gargarizans*. Although detected by both Zhang and Kumar's test and the divergence plotting, there was only one detected parallel site between the two species, and therefore we consider it suggestive rather than conclusive. Molecular parallel and convergent evolution often signals adaptive evolution (e.g. Castoe *et al.*, 2009; Liu *et al.*, 2014). Although we did not detect any signals of positive selection for any high-elevation species, it is functionally conceivable that the *HSP90AA1* gene may be involved in high-elevation adaptation. Strong radiation and large diurnal temperature variation are among the major environmental stressors at high-elevations (Cheviron and Brumfield, 2012). Considering that a primary function of the encoded protein is stabilizing proteins against heat stress, some modifications of the gene could be beneficial in drastically variable thermal regimes at high elevations.

	D	S	Q	N	R	N	K	L	S	E	L	L	R	Y	Y	T	S	A	S	G	D	E	M	V	S	L	K	D	Y	C	T	R	M	K
X. tropicalis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N. parkeri	.	N	.	.	.	**K**	.	.	.	M	.	.	.	**H**	.	.	.	.	.	.	.	.	T	**I**	.	.	**A**	.	.	**V**	**S**	.	.	.
Q. spinosa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R. kukunoris	.	T	.	.	.	**K**	.	.	.	.	.	.	.	**H**	.	.	.	.	.	.	.	.	.	**I**	.	.	**A**	.	.	**V**	**S**	.	.	.
R. chensiensis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.
O. margaretae	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B. gargarizans	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.	.
B. viridis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.	.
S. boulengeri	.	.	.	S	.	.	.	F	.	.	.	.	.	.	.	.	.	A	.	.	.	.	I	.	.	.	.	.	.	I	.	.	.	P
O. popei	.	.	.	.	.	R	.	.	.	.	.	.	.	Q	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 2 A section of the final alignment (amino-acid residues 169-202). Sites 174, 183, 192, 195, 198, 199 (bold) are identified as parallel evolution between two high-elevation species, *Nanorana parkeri* and *Rana kukunoris*.
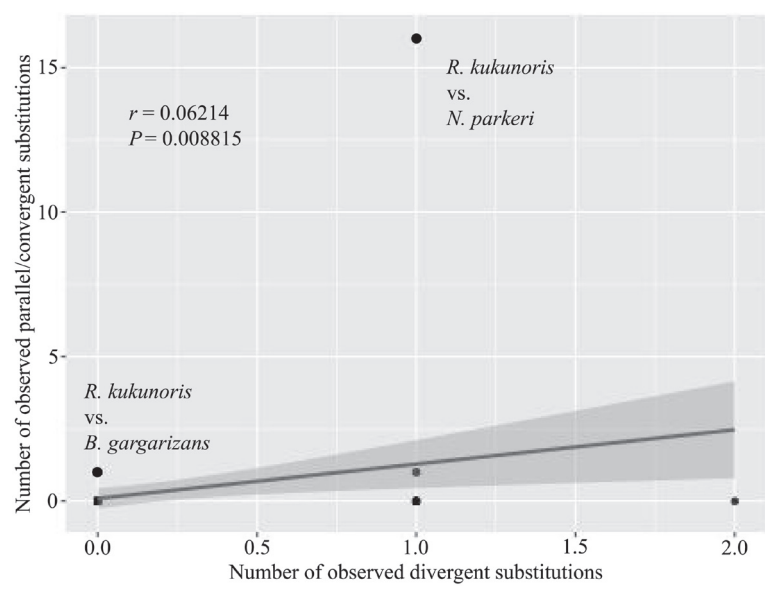
$r = 0.06214$
$P = 0.008815$

R. kukunoris vs. *N. parkeri*

R. kukunoris vs. *B. gargarizans*

Number of observed parallel/convergent substitutions

Number of observed divergent substitutions

Figure 3 Correlation between numbers of divergent substitution and numbers of convergent/parallel substitutions between all pairs of the ten amphibian species. The grey area represents 95% confidence interval. Two outliers show excessive numbers of parallel substitutions.

Further evaluation of how these parallel substitutions alter the physicochemical properties of the protein may shed some light on the potential adaptive values of these parallel substitutions.

Different methods clearly have various levels of sensitivity. Zhang and Kumar's test appeared to be most sensitive and detected multiple occasions of parallel evolution. On the other end of the spectrum, the gene tree method is understandably the most conservative. The parallel and convergent evolution has to be strong enough to overpower the phylogenetic signal in the data to draw two distantly related species together. Due to its simplicity and sensitivity nature, we recommend using Zhang and Kumar's test to screen the data first, and then using other methods to further elaborate the revealed patterns.

There are potential alternative explanations of the observed molecular parallel evolution of the *HSP90AA1* gene. The majority of our data were obtained through transcriptome sequencing. Alternative splicing or other process during the transcription process may alter the genomic data and produce erroneous results. Also, *HSP90* is a large gene family and there are several paralogous copies of the *HSP90AA1* gene. Inclusion of non-orthologous genes, or new and unknown isoform could produce the observed parallel evolution pattern.

Our study added one new case to a short list of known cases of molecular parallel and convergent evolution. Whether parallel and convergence evolution at molecular level is common remains a hotly debated topic. For example, Parker *et al.* (2013) argued a widespread molecular convergent and parallel evolution between echolocating bats and whales, but several others disputed their discovery (e.g. Thomas and Hahn, 2015; Zou and Zhang, 2015a). With the accumulation of new cases, we hope that this and other related controversies will be finally resolved.

References

Castoe T. A., de Koning A. P., Kim H. M., Gu W., Noonan B. P., Naylor G., Jiang Z. J., Parkinsons C. L., Pollock D. D. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci USA, 106: 8986–8991

Cheviron Z. A., Brumfield R. T. 2012. Genomic insights into adaptation to high-altitude environments. Heredity, 108: 354–361

Duellman W. E., Trueb L. 1994. Biology of Amphibians. Baltimore, USA: Johns Hopkins University Press

Gerchen J. F., Reichert S. J., Röhr J. T., Dieterich C., Kloas W., Stöck M. 2016. A single transcriptome of a green toad (*Bufo viridis*) yields candidate genes for sex determination and differentiation and non-anonymous population genetic markers. PLoS ONE, 11: e0156419

Hellsten U., Harland R. M., Gilchrist M. J., and 48 other authors. 2010. The genome of the western clawed frog *Xenopus tropicalis*. Science, 328: 633–636

Li P., Zhao E., Dong B. 2010. Amphibians and Reptiles of Tibet. Beijing, China: Science Press

Li Y., Liu Z., Shi P., Zhang J. 2010. The hearing gene prestin unites echolocating bats and whales. Curr Biol, 20: R55–R56

Liu Z., Qi F. Y., Zhou X., Ren H. Q., Shi P. 2014. Parallel sites implicate functional convergence of the hearing gene prestin among echolocating mammals. Mol Biol Evol, 31: 2415–2424

NCBI. 2018. https://www.ncbi.nlm.nih.gov/gene/3320

Parker J., Tsagkogeorga G., Cotton J. A., Liu Y., Provero P., Stupka E., Rossiter S. J. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature, 502: 228–231

Pyron R., Wiens J. J. 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol, 61: 543–583

Qiao L., Yang W., Fu J., Song Z. 2013. Transcriptome profile of the green odorous frog (*Odorrana margaratea*). PLoS ONE, 8: e75211

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rokas A., Carroll S. B. 2008. Frequent and widespread parallel evolution of protein sequences. Mol Biol Evol, 25: 1943–1953

Simonson T. S., Yang Y., Huff C. D., Yun H., Qin G., Witherspoon D. J., Bai Z., Lorenzo F. R., Xing J., Jorde L.

B. 2010. Genetic evidence for high-altitude adaptation in Tibet. Science, 329: 72–75.

Scott G. R., Schulte P. M., Egginton S., Scott A. L. M., Richards J. G., Milsom W. K. 2011. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. Mol Biol Evol, 28: 351–363.

Shen Y., Liu J., Irwin D. M., Zhang Y. P. 2010. Parallel and convergent evolution of the dim-light vision gene *RH1* in bats (Order: Chiroptera). PLoS ONE, 5: e8838

Shen Y., Liang L., Li G., Murphy R. W., Zhang Y. P. 2012. Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genet, 8: e1002788

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30: 1312–1213.

Stern D. L. 2013. The genetic causes of convergent evolution. Nat Rev Genet, 14: 751–764

Sun Y. B., Xiong Z. J., Xiang X. Y., Liu S. P., Zhou W. W., Tu X. L., Zhong L., Wang L., Wu D. D., Zhang B. L., Zhu C. L., Yang M. M., Chen H. M., Li F., Zhou L., Feng S. H., Huang C., Zhang G. J., Irwin D., Hillis D. M., Murphy R. W., Yang H. M., Che J., Wang J., Zhang Y. P. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proc Natl Acad Sci USA, 112: E1257–E1262

Thomas G. W, Hahn M. W. 2015. Determining the null model for detecting adaptive convergence from genomic data: A case study using echolocating mammals. Mol Biol Evol, 32: 1232–1236

Yang W., Qi Y., Bi K., Fu J. 2012. Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: A comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. BMC Genomics, 13: 588

Yang W., Qi Y., Fu J. 2016. Genetic signals of high-altitude adaptation in amphibians: A comparative transcriptome analysis. BMC Genet, 17: 134

Yang W., Lu B., Fu J. 2017. Molecular convergent evolution of the MYBPC2 gene among three high-elevation amphibian species. J Mol Evol, 84: 139–143

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol, 24: 1586–1591

Zhang J., Kumar S. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. Mol Biol Evol, 14: 527–536

Zhang J., Nielsen R., Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol, 22: 2472–2479

Zhao E., Yang D. 1997. Amphibians and Reptiles of the Hengduan Mountain region. Beijing, China: Science Press

Zou Z., Zhang J. 2015a. No genome-wide protein sequence convergence for echolocation. Mol Biol Evol, 32: 1237–1241

Zou Z., Zhang J. 2015b. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? Mol Biol Evol, 32: 2085–2096